The impact of excluding or including Death Certificate Initiated (DCI) cases on estimated cancer survival: A simulation study

Andersson, T. M-L., Myklebust, T. Å., Rutherford, M. J., Møller, B., Soerjomataram, I., Arnold, M., Bray, F., Parkin, D. M., Sasieni, P., Bucher, O., De, P., Engholm, G., Gavin, A., Little, A., Porter, G., Ramanakumar, A. V., Saint-Jacques, N., Walsh, P. M., Woods, R. R., & Lambert, P. C. (2021). The impact of excluding or including Death Certificate Initiated (DCI) cases on estimated cancer survival: A simulation study. *Cancer epidemiology*, *71*(Pt A), 101881. https://doi.org/10.1016/j.canep.2020.101881

Published in:
Cancer epidemiology

Document Version:
Peer reviewed version

The impact of excluding or including Death Certificate Initiated (DCI) cases on estimated cancer survival: a simulation study

Therese M-L Andersson*[1], Tor Åge Myklebust[2,3], Mark J Rutherford[4,5], Bjørn Møller[2], Isabelle Soerjomataram[5], Melina Arnold[5], Freddie Bray[5], D Maxwell Parkin [5,6], Peter Sasieni[7], Oliver Bucher[8], Prithwish De[9], Gerda Engholm[10], Anna Gavin[11], Alana Little[12], Geoff Porter[13], Agnihotram V Ramanakumar[14], Nathalie Saint-Jacques[15], Paul M. Walsh[16], Ryan R Woods[17], Paul C Lambert[1,4]

1. Department of Medical epidemiology and Biostatistics, Karolinska Institutet, Stockholm Sweden.
2. Cancer Registry of Norway, Institute of Population-based Cancer Research, Oslo, Norway
3. Department of Research and Innovation, Møre and Romsdal Hospital Trust, Ålesund, Norway
4. Department of Health Sciences, University of Leicester, Leicester, United Kingdom.
5. Cancer Surveillance Section, International Agency for Research on Cancer (IARC/WHO), Lyon, France.
6. Nuffield Department of Population Health, University of Oxford, Oxford, United Kingdom
7. Faculty of Life Sciences & Medicine, School of Cancer & Pharmaceutical Sciences, Guys Cancer Centre, Guys Hospital, King's College London, London, United Kingdom.
8. Department of Epidemiology and Cancer Registry, CancerCare Manitoba, Winnipeg, MB, Canada
9. Analytics and Informatics, Ontario Health (Cancer Care Ontario), Toronto, Ontario, Canada
10. Surveillance and Pharmacoepidemiology, Danish Cancer Society Research Center, Copenhagen, Denmark

11. Northern Ireland Cancer Registry, Queen's University Belfast, Northern Ireland, United Kingdom.
12. Cancer Institute NSW, Alexandria NSW, Australia
13. Canadian Partnership Against Cancer, Toronto, Ontario, Canada.
14. Research-Institute, McGill University Health Center, Montreal, Quebec, Canada
15. Nova Scotia Health Authority Cancer Care Program, Registry & Analytics, Halifax, Nova Scotia, Canada
16. National Cancer Registry Ireland, Cork, Ireland
17. Cancer Control Research, BC Cancer, Vancouver, British Columbia, Canada

Corresponding author: Therese M-L Andersson. E-mail: therese.m-l.andersson@ki.se

Key words: Cancer registry; Death certificate initiated cases; Survival; Simulation study

Word count: Abstract 248, Main text 3308, Highlights 65.

Abbreviations: DCN – death certificate notified; DCI – death certificate initiated; DCO – death certificate only; ICBP – International Cancer Benchmarking Partnership; RR – relative risk; RSR – relative survival ratio

Highlights:

- This simulation study shows that including cases initiated through death certificates in the survival analysis of population-based registry data will downwardly bias relative survival estimates.
- Excluding cases initiated through death certificates will in most situations overestimate survival.
- The extent of the bias depends on how missed cases differ from those registered through other routine sources.
- Registries should report the DCI proportion alongside the DCO proportion.

45

Abstract

Background: Population-based cancer registries strive to cover all cancer cases diagnosed within the population, but some cases will always be missed and no register is 100% complete. Many cancer registries use death certificates to identify additional cases not captured through other routine sources, to hopefully add a large proportion of the missed cases. Cases notified through this route, who would not have been captured without death certificate information, are referred to as death certificate initiated (DCI) cases. Inclusion of DCI cases in cancer registries increases completeness and is important for estimating cancer incidence. However, inclusion of DCI cases will generally lead to biased estimates of cancer survival, but the same is often also true if excluding DCI cases. Missed cases are probably not a random sample of all cancer cases, but rather cases with poor prognosis. Further, DCI cases have poorer prognosis than missed cases in general, since they have all died with cancer mentioned on the death certificates.

Methods: We performed a simulation study to estimate the impact of including or excluding DCI cases on cancer survival estimates, under different scenarios.

Results: We demonstrated that including DCI cases underestimates survival. The exclusion of DCI cases gives unbiased survival estimates if missed cases are a random sample of all cancer cases, while survival is overestimated if these have poorer prognosis.

Conclusion: In our most extreme scenarios, with 25% of cases initially missed, the usual practice of including DCI cases underestimated 5-year survival by at most 3 percentage points.

1. Introduction

Cancer survival, when estimated from population-based cancer registry data, is an important measure of the overall effectiveness of health systems given it estimates the average prognosis of cancer patients in the entire population. When comparing population-based cancer survival estimates between countries or jurisdictions, there has been some debate on how differences in registration processes and practices affect the observed survival differences (1). Previous studies have investigated different aspects, including the impact of: i) a failure to link cancer cases to their death information; ii) missing long-term survivors; iii) cancer cases notified from death certificates and iv) finding a date of recurrence instead of a date of diagnosis (2-5).

In this paper we focus on the impact on estimated survival of including or excluding cases notified through death certificates. Many cancer registries periodically receive notifications of cancer diagnoses based on death certificates, usually denoted as death certificate notified (DCN) cases (6, 7). For a majority of these DCN cases, the registry will also receive a notification from another source (e.g. pathology or hospital records). Yet for some cases, no additional notifications will be received, indicating these cases would not have been known to the registry were it not for the use of death certificate information. These cases are therefore not reported to the cancer registry when diagnosed.

For the DCN cases with no other notification to the registry, trace-back is often performed to actively ascertain when the cancer was first diagnosed and to verify that the case was a reportable cancer. The subset of DCN cases deemed reportable are referred to as DCI (death certificate initiated) cases (6, 7). DCI cases are therefore cases that are included in the cancer register solely

due to the use of death certificate notification, and would not have been reported from another source. DCI cases can be further subdivided into cases where trace-back was successful in finding a date of diagnosis and cases where the trace-back did not yield any additional information. The latter cases are commonly referred to as death certificate only (DCO) cases, and they are a subset of the DCI cases (6, 7). Some registries receive death certificate information more rapidly than notifications through other routine sources and therefore have a large group of cases initially notified from death certificates. However, these cases should not be referred to as DCI cases since they are reported to the registry through independent routine sources although at a later time. Only cases that would not have been known to the registry, if it was not for the death certificate, are DCI cases.

While it is important for cancer registries to include DCI cases to increase the completeness of cancer incidence statistics, including DCI cases when estimating survival will generally lead to biased results. The existence of DCI cases indicates that there are cases in the population who are not notified to the registry through the course of their disease and who are either alive, or have died without cancer mentioned as a cause of death. This is illustrated in Figure 1, the interest is in the survival of all cancer cases, i.e. the yellow box. However, some cancer cases are not registered through routine sources, and missed by the registry at diagnosis, represented by the grey solid box in Figure 1. A cancer registry that does not perform trace-back only includes the cases in the green solid box, those that are registered through routine sources. Some of these individuals will be alive at the time the cancer registry performs the survival analysis, some will have died with cancer mentioned on the death certificate and some will have died due to other causes, but all these cases are included. In the unlikely situation that these cases are a random sample of those in the yellow box this should yield unbiased estimates of survival. When a cancer

registry receives DCN cases, performs trace-back and then include the DCI cases, a subset of the missed cases are also included (the box with light green borders), the subset who died due to cancer (or where cancer is mentioned on the death certificate). The cases missed (not notified through routine sources, the solid grey box) that are still alive or died without cancer mentioned on the death certificate will not be retrieved, and continue to be missed by the registry. Since the DCI cases are not a random sample of the cases missed (solid grey box) by the registry, the inclusion of DCI cases when estimating survival can give biased results, even if the whole group of missed cases are a random sample of all cancer cases. The problem can be illustrated in a simple way by considering all cause survival among 1000 individuals. If the survival probability at 5 years is 0.8 and there is no censoring, one would expect there to be (800 people alive at 5-years (800/1000=0.8). If 20% of cases were initially missed (at random) then there would be 800 individuals initially with 800*0.8=640 alive at 5 years (640/800=0.8). Of those missed, one would expect 200*0.2=40 to die. Including these in the analysis leads to a 5-year survival of (640/(800+40) =76.2%, i.e. an underestimate as we have only added individuals to the denominator. There is often concern with respect to the validity of data from those registries unable to use death certificates to find additional cases, since it is known that excluding DCI cases will usually overestimate survival. However, the converse - that including DCI cases almost always underestimates survival is often not recognised.

The International Cancer Benchmarking Partnership (ICBP) SURVMARK-2 study aims to quantify disparities in cancer survival across high-income countries and identify possible reasons for them. As part of this international partnership, we performed a simulation study using a range of scenarios to quantify the impact on estimated cancer survival of including or excluding DCI

cases. The overarching aim was to comprehensively understand the potential impact of this bias on benchmarking cancer survival across populations.

2. Methods

To investigate the impact of including or excluding DCI cases on survival estimates, we simulated cohorts of 5000 cancer patients. For each cancer patient, a time of death due to cancer and a time of death due to other causes was simulated (8), and for each individual, their cause of death was determined by the event that occurred first: either death due to cancer, or death due to other causes. All survival times were censored at 10 years. We used three separate Weibull distributions for simulating time to death, representing a cancer site associated with low (Weibull parameter λ=0.61 and γ=0.63), medium (λ=0.4 and γ=0.6) and high (λ=0.12 and γ=0.64) cancer-specific survival, since the bias we wish to investigate can depend on the underlying cancer survival. We also used two levels (high and low, roughly corresponding to the survival of a 65 and an 80 year old in UK) of other cause (expected) survival, also with Weibull distributions (λ=0.034; γ=1.25 and λ=0.13; γ=1.19, respectively), since this can have an additional impact on the bias. The survival and hazard functions for both cancer-specific and other cause survival are shown in the Appendix Figure A1.

2.1 Simulating randomly missed cases

We simulated the proportion of the cancer cases who were missed, i.e. not notified to the registry, except possibly from death certificates, first assuming that these were a random sample of all cases. Three levels of missingness were investigated: 5%, 15% and 25%. This gave a total of 18 simulated scenarios: 3 levels of cause-specific survival, 2 levels of other cause survival and 3 levels for proportions of cases not reported to the registry, as listed in Table 1. Within each

simulated scenario, cases who were simulated to be missed by the registry and who died due to cancer within 10 years from diagnosis were classified as DCI cases. For simplicity, we assumed that the trace-back procedure found the correct date of diagnosis for all DCI cases, and hence there were no DCO cases. In actual registry data, DCO cases will exist, and they are usually excluded from survival analysis since their survival time is not known. This might have implications for the extent of bias in our simulations, however the direction of the bias is not altered.

2.2 Simulating non-randomly missed cases

We added another layer to the 18 base scenarios to investigate the impact of including a prognostic factor for death that is related to the extent of missingness. This prognostic factor was represented by a binary variable X (e.g. advanced stage), that affected the time to death due to cancer. The effect of Factor X was assigned a hazard ratio (HR) of 4, meaning that patients with the prognostic factor had a four times higher cancer-specific mortality rate than patients who did not have Factor X. Assuming 25% of the patients had this prognostic factor, we then simulated the 18 base scenarios as described above, where the probability of being missed differed by Factor X, while keeping the same overall probabilities of being missing. For each of the 18 main scenarios, 4 sub-scenarios (a, b, c and d) were simulated where the probability of being missed differed between those with and without Factor X with a relative risk (RR) of 1.5, 2, 3 and 5. For example, a RR of 1.5 means that those with Factor X were 50% more likely to be missed as those without the factor. The probability of being missed with and without Factor X, as represented in each scenario, is presented in Table 2. When simulating the time to death due to cancer in all these scenarios, the value of factor X for each individual was replaced by the value minus 0.25, so that the average hazard rate follows the Weibull distributions described above.

184

185 2.3 Estimating bias in cancer survival estimates

186 . We estimated relative survival ratios (RSR) at 1 and 5 years after diagnosis as measures of
187 cancer survival under two situations: (1) all missed cases were excluded from the analysis, thus
188 representing a situation where DCI cases are not included and (2) DCI cases are included. The
189 relative survival was estimated using flexible parametric models (9-11) with 4 degrees of
190 freedom, without inclusion of any covariates, and using the rate as specified from the Weibull
191 distribution used in simulation of time from death due to other causes for the expected mortality.
192 To calculate the bias in the RSR estimates introduced by excluding or including DCI cases, the
193 RSR estimates for situations (1) and (2) were compared with the true cancer specific survival
194 based on the Weibull distributions used for the simulations.. Both the absolute (as percentage
195 points) and relative (percentage) differences were calculated. The proportion of DCI cases was
196 also estimated as the difference in the number of cases included for the two situations, divided by
197 the number of cases included for situation (2). All results presented are averages based on 1000
198 simulations for each scenario.

199

200 2.4 Sensitivity analysis

201 Scenarios with HRs for Factor X of 1.5 and 2 were also simulated, and results from those
202 simulations are provided in the Appendix.

203

204 **3. Results**

205 3.1 Randomly missed cases

206 When cases who are missed by the registry were a random sample of all cancer cases occurring in
207 the population, unbiased estimates for the RSR were obtained when DCI cases were excluded

(Figure 2). Including DCI cases however underestimated survival, since the DCI cases are a selection of those missed who have a poorer prognosis. The size of the bias introduced differed across the 18 simulated scenarios, with the most important factor being the proportion of cases missed. When 5% of cases were missed (scenarios 1-6), the bias was small, less than 0.5 percentage points for 1-year survival and 0.6 percentage points for 5-year survival. When 15% of cases were missed (scenarios 7-12) the bias in 1-year survival was still lower than 1.5 percentage point, and just above 1.5 percentage points for 5-year survival. The largest bias – 2.5 percentage points for 1-year and 2.8 for 5-year survival – occurred when 25% of cases were not notified (scenarios 13-18).

There was no clear trend in the extent of bias in terms of the prognosis of the cancer (low, medium or high survival), or the level of other cause survival. Rather it was the combination of cancer and other cause survival which was important, since the extent of bias depends on the proportion of the missed cases who were added when the DCI cases were included in the analysis. As the bias will also depend on the true RSR, the relative bias is also presented in Figure 2.

3.2 Non-randomly missed cases

For the next set of results (Figure 3) we assumed that cases with a poorer prognosis were more likely to be missed. In this analysis, the exclusion of DCI cases led to an overestimation of survival, and for many scenarios this overestimation was greater than the underestimation introduced when DCI cases were included. The bias introduced by either including or excluding DCI cases was largest for the scenarios where 25% of cases were missed by the registry, suggesting that the proportion of cases missed was the most important driver of potential bias.

When DCI cases were excluded, the gap between the estimated and true survival widened, with an increasing RR of being missed for those with Factor X. The opposite was true when DCI cases were included. The largest bias observed when including DCI cases was an underestimation of 1-year survival by 2.7 percentage points and 5-year survival by 2.9 percentage points. The largest bias observed when excluding the DCI cases was an overestimation of 1-year survival by 5.9 percentage points and 5-year survival by 5.4 percentage points. Again, there was no clear trend in the extent of bias in terms of cancer-specific survival, or other cause survival.

3.3 Proportion of Death Certificate Initiated cases

The proportions of DCI cases for each simulated scenario are presented in Table 3. The proportion of DCI cases depends on the proportion of missed cases, since it can never be higher than the proportion missed. For any given value of the proportion missed, the proportion of DCI cases decreased with increasing cause-specific survival, as there would be a diminishing number of cases who die from cancer. On the other hand, the proportion of DCI cases was higher for higher other cause survival. This is because a larger proportion of cases will die due to cancer if fewer die due to other causes. Finally, the proportion of DCI cases also increased with increasing RR of being missed for those with Factor X compared to those without Factor X.

3.4 Sensitivity analysis

For scenarios where the HR for Factor X was changed to 1.5 or 2, the pattern of the results were similar to the scenarios where the HR was 4, however, the bias introduced by excluding DCI cases was smaller with a lower HR (Appendix Figures A2 and A3). The bias introduced by including DCI cases was less affected by the size of the HR for Factor X.

4. Discussion

Our simulation study shows that performing trace-back to include DCI cases, does not resolve the problem of missing cases biasing survival estimates, and can in certain circumstances lead to an even larger bias than that resulting from excluding DCI cases from the analyses. The inclusion of DCI cases in cancer registries is a necessary procedure to achieve the highest possible completeness in terms of cancer incidence. When estimating survival, the inclusion of DCI cases will underestimate survival, while their exclusion will overestimate survival. The utilization of death certificates as a source for cancer notifications implies that some cancer cases are not reported to the registry when diagnosed, and even if those missed are a random sample of cases, inclusion of the DCI cases will lead to biased survival estimates. Thus, excluding DCI cases when estimating survival will lead to unbiased survival estimates only if those cases not notified represent a random sample of all cancer cases – which is unlikely in most situations– otherwise, survival will be overestimated if the missed cases have more severe disease.

In our study we have demonstrated the impact on survival estimates of including and excluding DCI cases. This has consequences for survival benchmarking. For two countries where one includes DCI cases that were successfully traced back and the other does not, both estimates of cancer survival will be biased, but in opposite directions. Even when comparing two populations with the same practice in terms of including or excluding DCI cases, the bias may be of different magnitudes depending on the true proportion missing within each registry, the mechanisms that dictate the degree of missingness and the amount of trace-back. The inclusion of DCI cases could also lead to greater underestimation if the trace-back doesn't find the true date of diagnosis but rather a later date such as that at recurrence, but this was not evaluated in this study. An

additional issue when investigating trends over calendar time is that there is less opportunity for cases diagnosed (and missed) more recently, to be obtained from death certificates due to their shorter follow-up.

All cancer registries participating in ICBP SURVMARK-2 include DCI cases, although the proportion of DCI cases is often unknown. Unfortunately, most cancer registries are not able to retrospectively identify DCI cases in their data as typically this information is superseded when other information relating to time prior to death is retrieved. However, for registries within ICBP SURVMARK-2 where the proportion of DCI cases is available, a proportion of about 15% can be observed for cancer sites with poor prognosis, indicating that scenarios 7 and 8 are plausible scenarios for a poor prognosis cancer. For cancer sites with better prognosis, a proportion of DCI cases of about 3-4% has been observed in real data, indicating that scenarios 3-6 are plausible. However, given the small number of registries that have information on DCI cases, and the uncertainty in the proportion of missed cases, we explored a wider range of scenarios in this study.

A few limitations should be noted in relation to our study. We did not simulate an age distribution within the data, but rather investigated two levels of other cause survival. In all simulations we assumed that the prognostic Factor X was only associated with cancer-specific survival, but not other cause survival, which might be violated if the prognostic factor is, for example, the presence of comorbidity. We also assumed that cause of death is recorded accurately for all cases. Another aspect that could be of interest is specification of DCO cases. In our simulations we assumed that the true date of diagnosis is found for all DCI cases, resulting in no DCO cases. We also assumed that all death certificates had been retrieved by the registry by

the time the survival analysis was performed, so all cases were correctly classified. Even so, this simulation study showed clearly how the inclusion of DCI cases underestimates survival, and excluding DCI cases instead overestimates survival if cases who were not notified were not a random sample of all cancer patients in the population.

The extent of bias largely depends on the proportion of cases who are not notified, but the bias also differs depending on the extent to which the missed cases are notified as DCI cases (i.e. the proportion of the missed cases who have died and had cancer mentioned on their death certificates). It is reassuring to see that our scenarios give a bias of at most 3 percentage points in the situation when DCI cases are included. It is by definition impossible to know the true proportion of cases missed by a registry, but the proportion of DCI cases serves as an important indicator in this respect. Registries should therefore report the proportion of DCI cases along with the more commonly reported proportion of DCO cases.

Conflict of interest

The authors declare no competing interests.

Funding

The ICBP is funded by the Canadian Partnership Against Cancer; Cancer Council Victoria; Cancer Institute New South Wales; Cancer Research UK; Danish Cancer Society; National Cancer Registry Ireland; The Cancer Society of New Zealand; NHS England; Norwegian Cancer Society; Public Health Agency Northern Ireland on behalf of the Northern Ireland Cancer Registry; DG Health and Social Care, Scottish Government; Western Australia Department of Health; Public Health Wales NHS Trust.

Acknowledgements

The authors would also like to thank Lucie Hooper, Samantha Harrison, Charles Norell, Shanta Keshwala and Charlotte Lynch of Cancer Research UK for managing the programme. The ICBP Clinical Committees for their advice. The ICBP SurvMark-2 Academic Reference Group for providing independent peer review and advice for the study protocol and analysis plan development. Finally, we are thankful to the ICBP Programme Board for their oversight and direction.

Author statement

Where authors are identified as personnel of the International Agency for Research on Cancer/WHO, the authors alone are responsible for the views expressed in this article and they do not necessarily represent the decisions, policy or views of the International Agency for Research on Cancer/WHO.

References

1. Beral V, Peto R. UK cancer survival statistics. Bmj. 2010;341:c4112.
2. Robinson D, Sankila R, Hakulinen T, Moller H. Interpreting international comparisons of cancer survival: the effects of incomplete registration and the presence of death certificate only cases on survival estimates. Eur J Cancer. 2007;43(5):909-13.
3. Moller H, Richards S, Hanchett N, Riaz SP, Luchtenborg M, Holmberg L, et al. Completeness of case ascertainment and survival time error in English cancer registries: impact on 1-year survival estimates. Br J Cancer. 2011;105(1):170-6.
4. Woods LM, Coleman MP, Lawrence G, Rashbass J, Berrino F, Rachet B. Evidence against the proposition that "UK cancer survival statistics are misleading": simulation study with National Cancer Registry data. Bmj. 2011;342:d3399.
5. Rutherford MJ, Moller H, Lambert PC. A comprehensive assessment of the impact of errors in the cancer registration process on 1- and 5-year relative survival estimates. Br J Cancer. 2013;108(3):691-8.
6. Parkin DM, Bray F. Evaluation of data quality in the cancer registry: principles and methods Part II. Completeness. Eur J Cancer. 2009;45(5):756-64.
7. Bray F, Parkin DM. Evaluation of data quality in the cancer registry: principles and methods. Part I: comparability, validity and timeliness. Eur J Cancer. 2009;45(5):747-55.
8. Crowther MJ, Lambert PC. Simulating biologically plausible complex survival data. Stat Med. 2013;32(23):4118-34.
9. Royston P, Parmar MK. Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. Stat Med. 2002;21(15):2175-97.
10. Nelson CP, Lambert PC, Squire IB, Jones DR. Flexible parametric models for relative survival, with application in coronary heart disease. Stat Med. 2007;26(30):5486-98.
11. Lambert PC, Royston P. Further development of flexible parametric models for survival analysis. Stata J. 2009;9(2):265-90.

372 Table 1. Combinations of probability of cases being missed in the registry, cancer-specific

373 survival, and other cause (non-cancer) survival included in the 18 simulated main scenarios.

Scenario	Probability missed	Cancer-specific survival	Other cause survival
1	0.05	Low	Low
2	0.05	Low	High
3	0.05	Medium	Low
4	0.05	Medium	High
5	0.05	High	Low
6	0.05	High	High
7	0.15	Low	Low
8	0.15	Low	High
9	0.15	Medium	Low
10	0.15	Medium	High
11	0.15	High	Low
12	0.15	High	High
13	0.25	Low	Low
14	0.25	Low	High
15	0.25	Medium	Low
16	0.25	Medium	High
17	0.25	High	Low
18	0.25	High	High

374

375 Table 2. Probability of a case with and without prognostic Factor X being missed by the registry,

376 in four sub-scenarios* for each of the 18 base scenarios.

Scenarios	Sub-scenario	Probability missed among cases without Factor X	Probability missed among cases with Factor X
1-6	a	0.044	0.066
1-6	b	0.040	0.080
1-6	c	0.033	0.100
1-6	d	0.025	0.125
7-12	a	0.133	0.200
7-12	b	0.120	0.240
7-12	c	0.100	0.300
7-12	d	0.075	0.375
13-18	a	0.222	0.333
13-18	b	0.200	0.400
13-18	c	0.166	0.500
13-18	d	0.125	0.625

377 * Sub-scenarios a to d represent relative risk of being missed in the registry of 1.5; 2; 3 and 5,

378 respectively

379 Table 3. Proportion of Death Certificate Initiated (DCI) cases in each simulated scenario.

Scenario	% DCI	Scenario	% DCI	Scenario	% DCI
1	3.7	7	11.3	13	19.4
1a	3.7	7a	11.3	13a	19.4
1b	3.8	7b	11.6	13b	19.8
1c	3.9	7c	12.0	13c	20.5
1d	4.1	7d	12.5	13d	21.3
2	4.4	8	13.4	14	22.6
2a	4.3	8a	13.2	14a	22.3
2b	4.4	8b	13.3	14b	22.5
2c	4.5	8c	13.6	14c	22.9
2d	4.6	8d	13.8	14d	23.3
3	3.0	9	9.3	15	16.2
3a	3.1	9a	9.6	15a	16.7
3b	3.2	9b	10.0	15b	17.3
3c	3.4	9c	10.6	15c	18.2
3d	3.6	9d	11.2	15d	19.3
4	3.9	10	11.9	16	20.3
4a	3.8	10a	11.8	16a	20.2
4b	3.9	10b	12.1	16b	20.6
4c	4.1	10c	12.5	16c	21.2
4d	4.2	10d	13.0	16d	21.9
5	1.4	11	4.4	17	8.0
5a	1.6	11a	5.2	17a	9.4
5b	1.7	11b	5.6	17b	10.1
5c	1.9	11c	6.1	17c	11.0
5d	2.1	11d	6.8	17d	12.2
6	2.2	12	7.0	18	12.4
6a	2.4	12a	7.6	18a	13.5
6b	2.5	12b	8.1	18b	14.2
6c	2.7	12c	8.6	18c	15.2
6d	3.0	12d	9.4	18d	16.4

380
381

382 Figure 1. Illustration of Death Certificate Initiated (DCI) cases as a subset of all cases of cancer
383 arising in the population.
384
385

Figure 2. Absolute and relative differences in 1- and 5-year relative survival ratios (RSR) for the 18 base scenarios* described in Table 1 (where the missed cases are a random sample of all cases): including or excluding death certificate initiated cases compared to the full cohort. Negative values refer to underestimation of survival, and positive values overestimation of survival.

Note that the absolute and relative differences are shown with different scales

* 5%, 15%, 25% missing registration for scenarios 1-6, 7-12 and 13-18 respectively with different combinations of low, medium and high cancer specific survival and level of other cause survival.

Figure 3. Absolute and relative differences in 1- and 5-year relative survival ratios (RSR) for the 72 simulation scenarios* described in Table 1 and Table 2: including or excluding death certificate initiated cases compared to the full cohort. For each of the 18 base scenarios, sub-scenario a to d are displayed with varying degrees of transparency, a with least and d with most transparent circles. Negative values refer to underestimation of survival, and positive values overestimation of survival.

Note that the absolute and relative differences are shown with different scales

*5%, 15%, 25% missing registration for scenarios 1-6, 7-12 and 13-18 respectively with different combinations of low, medium and high cancer specific survival and level of other cause survival. Sub scenarios a-d: Relative risk of being missed for those with Factor X (with higher risk of dying) relative to those without of 1.5, 2, 3 and 5, respectively.

Appendix

Figure A1. Cause-specific and other cause survival (a) and hazard (b) functions used in simulations, representing scenarios with low, medium and high cancer-specific survival and high and low other cause survival.

(a)

(b)

Figure A2: Absolute and relative differences in 1- and 5-year relative survival ratios (RSR) for the 72 simulation scenarios* described in Table 1 and Table 2 using hazard ratio of 1.5 for cases with the prognostic Factor X.

Note that the absolute and relative differences are shown with different scales

* 5%, 10%, 15% missing registration for scenarios 1-6, 7-12, 13-18 respectively with different combinations of low, medium and high cancer specific survival and level of other cause survival. Sub scenarios a-d: Relative risk of being missed for those with Factor X (with higher risk of dying) relative to those without of 1.5, 2, 3 and 5, respectively.

Figure A3. Absolute and relative differences in 1- and 5-year relative survival ratios (RSR) for the 72 simulation scenarios* described in Table 1 and Table 2 using hazard ratio of 2 for cases with the prognostic Factor X.

Note that the absolute and relative differences are shown with different scales

*5%, 15%, 25% missing registration for scenarios 1-6, 7-12, 13-18 respectively with different combinations of low, medium and high cancer specific survival and level of other cause survival. Sub scenarios a-d: Relative risk of being missed for those with Factor X (with higher risk of dying) relative to those without of 1.5, 2, 3 and 5, respectively.